



Semantics and knowledge organization

Hjørland, Birger

Published in:
Annual Review of Information Science and Technology

Publication date:
2007

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Hjørland, B. (2007). Semantics and knowledge organization. In *Annual Review of Information Science and Technology* (pp. 367-405). Information Today.

Semantics and Knowledge Organization

Birger Hjørland

Royal School of Library and Information Science, Copenhagen

Introduction: The Importance of Semantics for Information Science

The aim of this chapter is to demonstrate that semantic issues underlie all research questions within Library and Information Science (LIS, or, as hereafter, IS)¹ and, in particular, the subfield known as Knowledge Organization (KO). Further, it seeks to show that semantics is a field influenced by conflicting views and discusses why it is important to argue for the most fruitful one of these. Moreover, the chapter demonstrates that IS has not yet addressed semantic problems in systematic fashion and examines why the field is very fragmented and without a proper theoretical basis. The focus here is on broad interdisciplinary issues and the long-term perspective.

The theoretical problems involving semantics and concepts are very complicated. Therefore, this chapter starts by considering tools developed in KO for information retrieval (IR) as basically semantic tools. In this way, it establishes a specific IS focus on the relation between KO and semantics.

It is well known that thesauri consist of a selection of concepts supplemented with information about their semantic relations (such as generic relations or “associative relations”). Some words in thesauri are “preferred terms” (descriptors), whereas others are “lead-in terms.” The descriptors represent concepts. The difference between “a word” and “a concept” is that different words may have the same meaning and similar words may have different meanings, whereas one concept expresses one meaning.

For example, according to WordNet 2.1 (2005), the word “letter” has five senses, of which two are: (1) “a written message addressed to a person or organization” and (2) “a letter of the alphabet, alphabetic character.” In a thesaurus, these meanings are distinguished by, for example, parenthetical qualifiers, as in the *Thesaurus of ERIC Descriptors* (1987, p. 136):

Letters (Alphabet)

Letters (Correspondence)

The thesaurus manages synonymy relations by means of "Use/Used for" relations and homonymy relations by means of parenthetical qualifiers. Furthermore, by means of semantic relations between descriptors (concepts) such as narrower term (NT), broader term (BT), and related term (RT), the thesaurus establishes the structure of a subject field:

Most thesauri establish a controlled vocabulary, a standardized terminology, in which each concept is represented by one term, a descriptor, that is used in indexing and can thus be used with confidence in searching; in such a system the thesaurus must support the indexer in identifying all descriptors that should be assigned to a document in light of the questions that are likely to be asked. A good thesaurus provides, through its hierarchy augmented by associative relationships between concepts, a semantic road map for searchers and indexers and anybody else interested in an orderly grasp of a subject field. (Soergel, 1995, p. 369)

It should now be clear that a thesaurus is basically a semantic tool because the "road map" it provides is semantic: The relations between concepts that a thesaurus indicates are semantic relations.

What is the case with thesauri is more or less the case with all kinds of what Hodge (2000, online) has presented as knowledge organizing systems (KOS) in the following taxonomy:

Term Lists

- Authority Files
- Glossaries
- Dictionaries
- Gazetteers

Classifications and Categories

- Subject Headings
- Classification Schemes
- Taxonomies
- Categorization Schemes

Relationship Lists

- Thesauri
- Semantic Networks
- Ontologies

All these types of KOS represent selections of concepts more or less enriched with information about their semantic relations. Semantic networks, for example, are instances of KOS utilizing more varied kinds of semantic relations than thesauri do, whereas authority files are examples

of KOS displaying limited information about semantic relations. Because such systems are basically about concepts and semantic relations, knowledge about concepts and semantics is important for research into, and the use of, any of those systems. In other words, researchers in KO should ground their work in a fruitful theory of semantics. This kind of basic research has, however, been largely absent from IS.

Having argued that the various types of items which Hodge has identified as KOS may all be considered semantic tools, we will now take a closer look at the term “knowledge organizing systems.”

Hodge (2000) omits certain kinds of KOS—for example, bibliometric maps such as those provided by White and McCain (1998). In these maps, citation patterns may be generated by authors and/or by terms (e.g., from descriptors). Such maps thus display certain kinds of semantic relations on the basis of citing behavior (and the relation between terms on such a map suggests a certain kind of *semantic distance*). It is thus important to include bibliometrics within the concept of KOS for both theoretical and practical reasons.

There are other kinds of KOS that Hodge (2000) does not consider. It could be argued that encyclopedias, libraries, bibliographical databases, and many other concepts used within IS should be considered as KOS. Furthermore, concepts outside IS, such as the system of scientific disciplines or the social division of labor in society, also constitute very fundamental kinds of KOS. Indeed, KOS in a narrow, IS-oriented sense are those systems related specifically to organizing bibliographical records (in databases), whereas KOS in a wide, general sense are related to the organization of literatures, traditions, disciplines, and people in different cultures.

Although all the KOS listed by Hodge, as well as others, such as bibliometric maps, may be considered semantic tools, not all kinds of KOS can be identified as such. The system of scientific disciplines, for example, is not a semantic tool. The term “semantic tool” should be reserved for systems that provide selections of concepts more or less enriched with information about semantic relations; KOS should be used as a broader term including, but not limited to, semantic tools.

The field of KO within IS is thus concerned with the construction, use, and evaluation of semantic tools for IR. This insight brings semantics to the forefront of IS. This view is shared by Khoo and Na (2006, p. 207), who declare that “natural language processing and semantic relations, in particular, point the way forward for information retrieval in the 21st century.”

Because concepts provide the meaning behind words and semantics is the study of meaning, the study of concepts, meaning, and semantics should form one interdisciplinary subject field. However, the relevant literature is very scattered and difficult to synthesize, for it covers, among other fields, philosophy, linguistics, psychology and cognitive science, sociology, computer science, and information science. In addition to the disciplinary scattering of research in semantics, the field is based on

different epistemological assumptions whose roots extend back hundreds of years into the history of philosophy. Moreover, the field seems theoretically muddled.

Semantics, by the way, is not concerned solely with word meaning. Pictures as well as other signs are also the objects of semantics. The way semantics is viewed and discussed in this chapter may seem, in the eyes of many people, more like semiotics (the study of signs in general) than semantics as commonly understood. The relation between semantics and semiotics is itself a controversial issue. The focus on semantics rather than semiotics in this chapter is motivated by the fact that the-saural relations (like KOS in general) are semantic relations.

The Status of Semantic Research in Information Science

Van Rijsbergen (1986, p. 194) has pointed out that the concept of meaning has been overlooked in IS and discussed why the whole area is in crisis. The fundamental basis of all the previous work—including his own—is wrong, he claims, because it has been based on the assumption that a formal notion of meaning is not required to solve IR problems. This statement by a leading researcher should encourage closer cooperation between IS and other fields conducting research in semantics. Few researchers have, however, risen to the challenge and not much consideration has been given to the nature of semantics and its implications for IS.

Some of those addressing semantic issues in KO and IS are Bean and Green (2001); Beghtol (1986); Blair (1990, 2003); Bonnevie (2001); Brooks (1995, 1998); Budd (2004); Dahlberg (1978, 1995); Daily (1979); Doerr (2001); Foscett (1977); Frohmann (1983); Green, Bean, and Myaeng (2002); Hammwöhner and Kuhlen (1994); Hedlund, Pirkola, and Kalervo (2001); Hjørland (1997, 1998); Khoo and Na (2006); Qin (1999, 2000); Read (1973); Song and Galardi (2001); Stokolova (1976, 1977a, 1977b); and Vickery and Vickery (1987).

These contributions are very different and difficult to present in any coherent way because they are not related to each other or systematically related to broader views. Some of them try to base their view on an explicit philosophy (e.g., “Activity Theory” [Hjørland, 1997] or Wittgenstein’s philosophy [Blair, 1990, 2003; Frohmann, 1983]); others, for example, Vickery and Vickery (1987), base their view on cognitive psychology, but many simply present their own commonsense views without attempting to ground them in general theories (e.g., Foscett, 1977). A book such as that by Green, Bean, and Myaeng (2002) should be praised for its attempt to present an interdisciplinary perspective. Both this book and reviews such as Khoo and Na’s (2006) fail, however, to consider much previous research within IS (such as many of the references listed here) and thus lack a historical perspective on the relation between semantics and IS. They also fail to provide a discussion of basic issues in semantics or to argue systematically for a specific theoretical

view. This state of the art leaves us without a clear line of progress. Without proper theoretical frames of reference, empirical research becomes fragmented and almost impossible to perceive as a whole.

Much research is also based on technicalities and does not show much concern for basic semantic issues. This is the case with bibliometric research about semantic relationships among highly cited articles (e.g., Song & Galardi, 2001), with the technique known as "latent semantic indexing" or "latent semantic analysis" (Ding, 2005; Dumais, 2004) and, of course, with a new concept considered by many the most important frontier in KO, "the semantic Web" (Antoniou & van Harmelen, 2004; Berners-Lee, Hendler, & Lassila, 2001; Fensel, Hendler, Lieberman, & Wahlster, 2003). Some authors (e.g., Budd, 2004) have introduced important philosophical and semantic views into IS, but have not fully explored their implications for KO. There is a danger that the philosophical insights remain too isolated and vague.

The question concerning the relationship between semantics and KO may be turned upside down and we may ask from which theoretical perspectives KO has been approached. Which views of semantics have been implied by those approaches? KO has a long tradition within IS: Among the classics in the field is Bliss (1929). In order to discuss the relations between semantics and KO we should ask: What approaches have been used in the field of KO in the course of its history? How do they relate to semantic theory? Broughton, Hansson, Hjørland, and López-Huertas (2005) have suggested that the following traditions are the most important ones in KO:

1. The traditional approach to KOS expressed by classification systems used in libraries and databases, including the Dewey Decimal System (DDC), the Library of Congress Classification (LCC), and the Universal Decimal Classification (UDC)
2. The facet-analytical approach founded by Ranganathan
3. The IR tradition
4. User oriented/cognitive views
5. Bibliometric approaches
6. The domain-analytic approach
7. Other approaches, including semiotic, "critical-hermeneutical," discourse-analytic, and genre-based ones, as well as those that place emphasis on document representations, document typology and description, markup languages, document architectures, and so forth

Given that KOS essentially are semantic tools, should different approaches to KO reflect different approaches to semantics? This question can be answered only briefly here. The traditional approach to classification introduced the principle of literary warrant and thus located

semantic relations in the scientific and scholarly literature. This was (and is) often done on positivist premises: The scientific literature is seen as representing facts about knowledge and structures in knowledge, and subject specialists are deemed capable of making true and objective representations of it in KO (thus tending to neglect conflicting evidence and theories). The facet-analytic approach tends to base KO on a priori semantic relations. These are derived from the application of (logical) principles rather than from the study of evidence in literatures (although this latter approach, too, is visible to some degree within the facet-analytic tradition). The IR tradition sees semantic relations as statistical relations between signs and documents. It is atomistic in the sense that it does not consider how traditions, theories, and discourse communities have formed the very statistical patterns it observes. User-oriented and cognitive approaches tend to replace literary warrant with empirical user studies and thus to base semantic relations on users rather than on the scientific literature. The bibliometric approach considers documents to be semantically related if they cite each other, are co-cited, or are bibliographically coupled. Again, the semantic relations are based on some kind of literary warrant, but in a way quite different from that of the traditional approach. The domain-analytic approach is rather traditional in its identification of semantic relations based on literary warrant. However, it is not positivist, for it regards semantic relations as determined by theories and epistemologies, which more or less influence all fields of knowledge. Many recent approaches to KO, including semiotic and hermeneutic approaches may be considered to be related to the domain-analytic approach.

What this suggests is that different approaches to KO imply different views on semantics. This point, however, has not been previously considered in the literature.

Semantics and the Philosophy of Science

The different theories and epistemologies that are in competition with one another may be more or less fruitful (or harmful) for information science. It is important to realize this and to take the risk of defending a particular theory. If this is not done, other views will never be sufficiently falsified, confirmed, or clarified. In the process of defending a particular view, one learns what other views it is necessary to reject. As pragmatist philosophers have long suggested, in order to make our thoughts clear, we have to ask what practical consequences follow from taking one or another view (or meaning) as true. If our theory (or meaning) does not have any practical implications, then it is of no consequence.

Peregrin (2004) has suggested that there are two dominant paradigms in semantics: One elaborated by logical positivists such as Rudolph Carnap (and the young Wittgenstein) and another developed by pragmatist philosophers such as John Dewey, which also draws on the insights of the late Wittgenstein. Positivist semantics suggests that

expressions “stand for” entities and their meanings are the entities stood for by them. Pragmatist semantics suggests that expressions are tools for interaction and their meanings are their functions within the interaction, giving them the capacity to support it in their distinctive ways.² Hjørland and Nissen Pedersen (2005) have used this dichotomy to set the foundations of a theory of classification for IR. Their arguments may be summarized as follows:

1. Classification is the ordering of objects (or processes or ideas) into classes on the basis of some properties. (The same is the case when terms are defined: It is determined what objects fall under the term.)
2. The properties of objects are not just “given” but are available to us only on the basis of some descriptions and pre-understandings of those objects.
3. Description (or every other kind of representation) of objects is both a reflection of the thing described and of the subject creating the description. Descriptions are more or less purposeful and theory-laden. Pharmacologists, for example, in their description of chemicals, emphasize their medical effects, whereas “pure” chemists emphasize other aspects of the chemicals such as their structural properties.
4. The selection of the properties of the objects to be classified must reflect the purpose of the classification. There is no “neutral” or “objective” way to select properties for classification because any choice facilitates some kinds of use while limiting others.
5. The (false) belief that there exist objective criteria for classification may be termed “empiricism” or “positivism,” whereas the belief that classifications always reflect a purpose may be termed “pragmatism.”
6. Different domains (e.g., chemistry and pharmacology) may need different descriptions and classifications of objects to serve their specific purposes in the social division of labor in society. The criteria for classification are thus generally domain-specific. Different domains develop specific languages (languages for specific purposes, or LSPs) that are useful for describing, differentiating, and classifying objects in their respective domain.
7. In every domain, there exist different theories, approaches, interests, or “paradigms,” which also tend to describe and classify objects according to their respective views and goals.
8. Any given classification or definition will *always* be a reflection of a certain view or approach to the objects being classified. Ørom (2003), for example, has shown how different library classifications reflect different views of the arts. Ereshefsky (2000) has argued that Linnaean classification is based on criteria that are

pre-Darwinian and thus problematic. Sometimes, however, a given classification seems to be immune to criticism. This may be the case with the periodic table of elements in chemistry and physics. Such immunity is caused by a strong consensus in the underlying theory.

9. A given literature to be classified is always—to some extent—a merging of different domains and approaches/theories/views. Such different views may be explicit or implicit. If they are implicit, they can be uncovered by theoretical and philosophical analysis.
10. Classifications and semantic systems that do not consider the different goals and interests reflected in the literature of a given domain are “positivist.” The criteria for classification should be based on an understanding of the specific goals, values, and interests at play. They are not to be established a priori, but by “literary warrant”—i.e., by examining the literature. This cannot be done in either a “neutral” or an “objective” way, but can be accomplished by considering the different arguments.

In her reply, Sparck Jones (2005, p. 601) has acknowledged this pragmatic point of view. Her final suggestion is, however:

One of the most important techniques developed in retrieval research and very prominent in recent work, namely relevance feedback, raises a more fundamental question. This is whether classification in the conventional, explicit sense, is really needed for retrieval in many, or most, cases, or whether classification in the general (i.e., default) retrieval context has a quite other interpretation. Relevance feedback simply exploits term distribution information along with relevance judgements on viewed documents in order to modify queries. In doing this it is forming and using an implicit term classification for a particular user situation. As classification the process is indirect and minimal. It indeed depends on what properties are chosen as the basic data features, e.g., simple terms and, through weighting, on the values they can take; but beyond that it assumes very little from the point of view of classification. It is possible to argue that for at least the core retrieval requirement, giving a user more of what they like, it is fine. Yet it is certainly not a big deal as classification per se: in fact most of the mileage comes from weighting. And how large that mileage can be is what retrieval research in the many experiments done in the last decade have demonstrated, and web engines have taken on board.

I agree that meanings and classification criteria are implicit in the literature to be retrieved, as outlined here. Sparck Jones asks “whether classification in the conventional, explicit sense, is really needed for retrieval.” My answer to this question is that no retrieval mechanism (and also any definition of “relevance”) is ever neutral; it always considers some interests at the expense of others. To distinguish between such views is to make a kind of classification. To believe in a technical solution employing “relevance feedback” is to fall into the positivist trap. The vision of automated feedback and value-free systems is seductive but based on problematic philosophical assumptions.

This *ARIST* chapter espouses the pragmatist understanding of concepts, meaning, and semantics. This perspective may be able to address fundamental problems in KO and IR from a new and promising angle. The theoretical standpoint is that expressed by the American philosopher Hilary Putnam. He gives a résumé of his criticism in a paper bearing the apt title “The meaning of ‘meaning’”:

Traditional semantic theory leaves out only two contributions to the determination of extension—the contribution of society and the contribution of the real world! (Putnam, 1975, p. 164)

Putnam is also known as a philosopher in the pragmatist tradition. We may thus list three characteristics of his (and our) philosophical point of departure:

- A focus on the relation between meaning and the real world (realism)
- A focus on the functional/pragmatic nature of meaning (pragmatism)
- A focus on the development of meaning in a social context (historicism and meaning collectivism/holism)

We can say with Putnam that these principles have been very much ignored in semantic theory. We can also assert that they have also been ignored to a large extent in fields such as IS, despite the fact that, as shown here, these fields are heavily dependent on semantics.

Semantics and Subject Knowledge

Advanced semantic tools demand proper subject knowledge for their design and administration, as well as for their use and evaluation. This follows from the realist philosophical position formulated previously: Knowledge of semantic relations between terms requires world knowledge about the relations between the objects that the terms refer to. You cannot determine the semantic relations between the words

“Copenhagen” and “Denmark” unless you know that Copenhagen is a part of Denmark.

This has been well known in the world of research libraries and bibliographical databases as well as in education for librarianship. The Medline database, for example, demands that a “prospective indexer must have no less than a bachelor’s degree in a biomedical science, and should also have a reading knowledge of one or more modern foreign languages. An increasing number of recent recruits hold advanced degrees in biomedical sciences” (National Library of Medicine, 2005, online).

Concerning the construction of ontologies for gene technology, Bada, Stevens, Goble, Gil, Ashburner, Blake, et al. (2004, p. 237) write:

One of the factors that account for GO’s [Gene Ontology’s] success is that it originated from within the biological community rather than being created and subsequently imposed by external knowledge engineers. Terms were created by those who had expertise in the domain, thus avoiding the huge effort that would have been required for a computer scientist to learn and organize large amounts of biological functional information. This also led to general acceptance of the terminology and its organization within the community. This is not to say that there have been no disagreements among biologists over the conceptualization, and there is of course a protocol for arriving at a consensus when there is such a disagreement. However, a model of a domain is more likely to conform to the shared view of a community if the modellers are within or at least consult to a large degree with members of that community.

These quotations do not constitute a new view. Earlier, Richardson and Bliss had considered the implications of the need of subject knowledge for education in librarianship and IS:

Again from the standpoint of the higher education of librarians, the teaching of systems of classification ... would be perhaps better conducted by including courses in the systematic encyclopedia and methodology of all the sciences, that is to say, outlines which try to summarize the most recent results in the relation to one another in which they are now studied together. (Richardson, quoted in Bliss, 1935, p. 2)

Furthermore, at the close of her linguistic investigation into semantic relations, Murphy (2003, p. 242) draws the following conclusion:

Plainly, the topic of lexical semantic paradigms has not been exhausted, and the metalinguistic approach discussed in this book gives rise to a number of new directions for lexicological

research. It fits with (and exploits) a general trend in linguistic research to appreciate the particular relations that language engages in: the relation between language and context, language and conceptualization, language and linguistic behavior. While [Leonard] Bloomfield (1985/[1936]) argued that linguists should ignore meaning because it is not properly “linguistic,” to hold such a position in the current disciplinary context is untenable, since many if not most (if not all) linguistic phenomena cross boundaries between the linguistic, the conceptual, and the communicative. In the case of lexical relations, this means that those who study it are not just linguists, but metalinguists.

The domain-analytic view in information science is an attempt to provide subject knowledge within the boundaries of IS in a way that still makes it possible for professionals to have a clear identity as information scientists (cf. Hjørland, 2002a). Teaching librarians and information specialists the content of a paper such as that of Ørom (2003) would provide a better basis for all kinds of information work related to the arts. In addition, it would provide certain possibilities for generalization to other domains. In this way, information specialists would provide domain-specific knowledge while operating within a framework that allows IS to have a specific identity.

Domain knowledge is a problem not only for IS but also for linguistics and many metasciences (such as cognitive science and the sociology of science). Much cognitive and linguistic theory regarding concepts, meaning, and semantics is strongly constrained by attempts to avoid “world knowledge.” The importance of subject knowledge has theoretical implications for how concepts should be defined and semantic relations determined (whether by human or by machine). It has implications for answering the question: What kind of information is needed in order to determine the semantic relations between two terms A and B? This question is considered in the next section.

Semantics and Its “Warrant”

Theories of semantics should be formulated in ways that provide methodological implications for determining meanings and relations in semantic tools such as thesauri and semantic networks. Often such implications are not clear; this renders the theories vague and unhelpful. Murphy (2003, p. 111), for example, has observed:

From the WordNet literature available, it is often difficult to determine the bases on which design decisions in WordNet are made. For example, Miller (1998) notes that Chaffin, Herrmann, and Winston (1988) identified eight types of meronymy and Iris, Litowitz, and Evens (1988) distinguished

four types, but he does not indicate how it was determined that WordNet should distinguish only three types.

Similarly it is often unclear on what bases specific decisions are made in classification systems such as DDC or in thesauri such as the *Thesaurus of Psychological Index Terms* (Kinkade, 1974; Walker, 1997).

Frohmann (1983) has discussed the semantic bases and theoretical principles of some classification systems. His is one of the few papers in IS to recognize that problems in classification should be seen as problems related to semantic theories. He observes that concepts such as "dog," "cat," "whale," "pike," and "owl" may be grouped or classified in different ways:

For example, one principle of division divides the set according to nocturnal and diurnal characteristics. In this case, "cat" and "owl" belong to the first category, and the other terms to the second. Another principle of division separates mammals from non-mammals. In that case, "dog," "cat," and "whale" belong to the first category, whereas "pike" and "owl" belong to the second. Other divisions may be recognized (e.g., "land creatures," "water creatures," and "flying creatures"). (Frohmann, 1983, pp. 15–16)

Frohmann presents two semantic theories. The first holds that the categories to which a concept belongs are given a priori as part of the "meaning" of the term for that concept. According to the second, the categories to which a concept belongs must be found in the specific literature or discourse of which the associated term is a part. Consequently, the semantic relations are not given a priori, but are formulated a posteriori. This distinction has implications for classification theory. Frohmann demonstrates that Austin's PRECIS system (as an example) is based on a priori semantics and therefore open to an argument from Wittgenstein's later philosophy of language. According to Frohmann, KO systems cannot be both machine-compatible and adequate, as Austin claimed (although he does not rule out other ways to construe systems that are both machine-compatible and adequate).

Thus, a basic problem in KO is whether semantic relations are a priori or a posteriori: whether they can be known before examining the literature or only after such an examination has been carried out. What kind of literary warrant (or other kind of warrant) is needed in order to identify semantic relations and classify concepts?

This question is also related to one about the possibility of universal solutions to KO because a posteriori relations are unlikely to be universal. According to Frohmann (1983), the Classification Research Group (CRG) in England realized that semantic relations are a posteriori relations and have to be determined by examining specific disciplinary literatures individually. However, neither Frohmann

himself nor the literature from the CRG and the Bliss Bibliographical Classification goes into details about precisely *how* concepts should be defined and their relations identified. Although it is correct that the CRG (and the Bliss Classification System, 2nd ed.) work on the basis of examining specific literatures, it is not clear—at least to this author—to what extent semantic relations are taken from the literature to be classified or are imposed on that literature. My opinion is that those systems are based on a priori principles to a greater degree than Frohmann suggests. There is a tendency within the facet-analytic tradition to work with universal categories like time and space and to classify the literature in relation to such pre-established categories. I believe this will be clearer when we analyze different theories of concepts and semantics.

Let us look at some theoretical possibilities about the nature of concepts and semantic relations. These might be:

- Query/situation specific or idiosyncratic
- Universal, Platonic entities/relations
- “Deep semantics” common to all languages (or inherent in cognitive structures)
- Specific to specific empirical languages (e.g., Swedish)
- Domain- or discourse-specific
- Other (e.g., determined by a company or a workgroup, “user-oriented”)

Before discussing these possibilities separately, let us adumbrate some general considerations about the nature of semantic relations. Semantic relations are often displayed in standard lexica—for example, in the *Longman Synonym Dictionary* (1986), WordNet, and similar semantic tools. However, it is well known that, for example, synonyms are seldom synonyms in all contexts. It thus becomes important not to think of semantic relations as simply “given,” but to ask: When are two concepts A and B to be considered synonyms (or homonyms or otherwise semantically related)? When is a semantic relation? We should again ask the pragmatist question: What difference does it make whether, in a given situation, we choose to consider A and B as semantically related in a specific way? This may look strange, given that many semantic relations seem intuitively “given” or authoritatively established in standard dictionaries.³

This relativity of meaning is also evident from Ogden and Richards’s (1923) famous triangle of meaning (Figure 8.1).⁴

The triangle implies that the referent of an expression—that is, a word or another sign or symbol—is relative to different language users. As Peirce (1931–1958, Vol. 2, p. 228) put it:

A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object* [or referent]. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen.

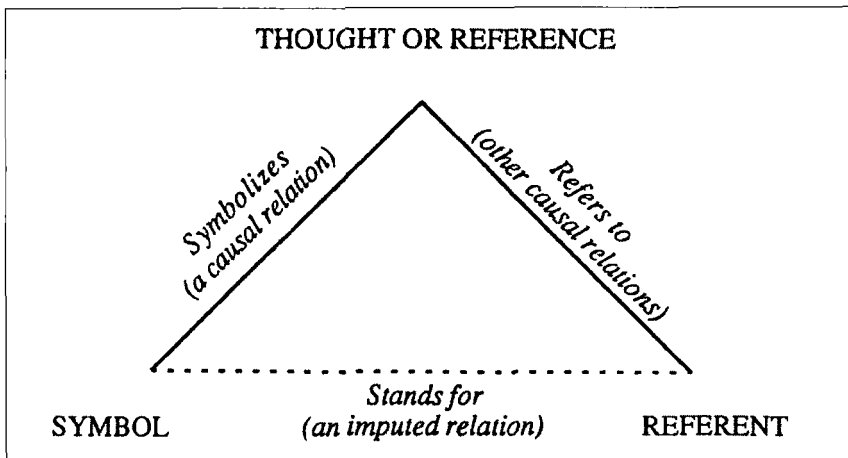


Figure 8.1 Ogden and Richards's (1923) semiotic triangle.

Concerning Query/Situation-Specific or Idiosyncratic Semantics

"I use a word," Humpty Dumpty said,
in rather a scornful tone,
"it means just what I choose it to mean—
neither more nor less."

"The question is," said Alice,
"whether you *can* make words mean
so many different things."

"The question is," said Humpty Dumpty
"which is to be master—that's all."
(Carroll, 1899)

It is important to keep in mind that concept determination and semantic relations are to be used in, for example, query expansion (automatic or manual) as well as in query precision and query formulation. In a way, it is the specific "information need" that determines which relations are

fruitful and which are not in a given search session. A semantic relation that increases recall and precision in a given search is relevant in that situation. Creative information searchers do just that: They provide search strategies that retrieve a fruitful set of documents by combining terms in unusual ways. Different terms may be combined using the Boolean operator OR in a given search. By implication, they are regarded as equivalent terms (or synonyms) in the situation, even though they are not normally considered synonyms. For example, antonyms and contrary terms are different from synonyms. Yet, in IR, it is often useful to conduct searches using antonyms because certain phenomena may be discussed in relation to their opposites. The implication is that, in a given search, it might be useful to regard antonyms as synonyms.

This pragmatist point of departure is important to keep in mind in developing a theory of concepts and semantics. Semantic relations relate to a given task or situation and not all users of a given set of semantic relations will share the same view of which terms are equivalent. On the other hand, it is clear that if we base a semantic theory on an individualistic/idiosyncratic view of concepts and semantics, it is not possible to design systems for more than one user or situation—an absurd conclusion. We need more stable principles on which to determine semantic relations. We need a semantic theory about the meaning of words as forms of *typified practices*. Knowledge about semantics in typified practices may then be used by information searchers in order to include or exclude certain documents.

Concerning Universal, Platonic Entities/Relations

Mathematicians are, probably more than other professionals, Platonists. They believe that the mathematical concepts such as π (pi) have always existed and had only to be discovered. π is semantically related to the “radius” and the “perimeter” of a “circle” (because it is defined as the relation between those concepts). This semantic relation is universal and *given* (although the symbols chosen are conventional). According to Platonism, the meaningfulness of a general term is constituted by its connection with an abstract entity, the (possibly) infinite extension of which is determined independently of our classificatory practices (cf. Haukioja, 2005).

The question for us is: Is it also *a priori* in the sense Frohmann (1983) meant? It may be sufficient to say that the semantics of, for example, mathematical concepts are not simply intuited by the individual indexer. They have to be determined by considering the mathematical literature (or by people educated in that literature). Even if the basic method of knowing in mathematics involves a kind of rational intuition, this does not imply that semantic relations in mathematics should be considered to be *a priori* in KO.

Concerning “Deep Semantics” Common to All Languages or Inherent in Cognitive Structures (A Priori Relations)

Much research on semantics is based on the assumption that concepts are somehow “hardwired” to our mind or brain, for example, in our so-called “mental lexicon.” This is perhaps most clearly seen in research on color concepts.

Berlin and Kay’s (1969) book *Basic Color Terms: Their Universality and Evolution* has had a major impact on how we view color terms. The authors argued for the universality and evolutionary development of eleven Basic Color Terms. Some salient characteristics of this universalist position have been summarized by one of its main critics, Barbara Saunders (1998, online):

The relation between Munsell, the workings of the visual system, and the colour naming behaviour of people, is so tight it can be taken to be a causative law. Diversity of colour-naming behavior is defined as a system-regulated stability evinced by Evolution. The full lexicalisation of the human colour space is designated Evolutionary Stage Seven, as in American English; languages below this level are the fossil record.

Berlin and Kay’s (1969) view of color concepts contrasts with a cultural-relative view, according to which our color concepts (and semantics in general) are determined primarily not by our visual system but by our need to act in relation to the colored environment:

Sociohistorical psychology emphasizes the fact that sensory information is selected, interpreted, and organized by a social consciousness. Perception is thus not reducible to, or explainable by, sensory mechanisms, per se. Sapir, Whorf, Vygotsky, and Luria do not deny the existence of sensory processes—they maintain that sensory processes are subordinated to and subsumed within ‘higher’ social psychological functions. (Ratner, 1989, p. 361)⁵

We may thus conclude that the universality of color terms is controversial. The dominant view is cognitivist and maintains the universality of concepts, but a well-argued minority maintains a relativist view of color concepts, a position related to the pragmatist standpoint.

A certain version of “deep semantics” is the theory of semantic primitives according to which every word can be broken up into primitive kernels of meaning, *semantemes* (also called *semantic features* or *semantic components*). Semantemes are terms that are used to explain other terms or concepts but cannot themselves be explained by other terms. The process of breaking words down into semantemes is known

as *componential analysis* and has been most often used to analyze kinship terms across languages. The components are often given in considerable detail. For instance, kinship terms like those shown in Table 8.1 might have three components: sex, generation, and lineage. Sex would be male or female; generation would be a number, with 0 = reference point's generation, -1 = previous generation, +1 = next generation; and lineage would be either direct, colineal (as in siblings), or ablineal (as in uncles and aunts).

Cruse (2001, p. 8758) has characterized the theory of semantic primitives as an "influential approach, much criticized but constantly reborn." He also writes (p. 8759)

In the earliest versions of componential analysis, the components were the meanings of words, and the aim of the analysis was to extract a basic vocabulary, in terms of which all non-basic meanings could be expressed. Generally speaking, the features recognized by earlier scholars had no pretensions to universality, and indeed were often avowedly language-specific. Later scholars aimed at uncovering universals of human cognition, a finite "alphabet of thought." Accessible introductions to componential analysis can be found in Nida (1975) and Wierzbicka (1996).

According to Sparck Jones (1992, p. 1609), this theory was influential in early thesaurus construction: "A thesaurus was seen as providing a set of domain-independent *semantic primitives*."

Theories about "innate ideas" (including concepts and semantic relations) have roots far back in the history of philosophy and are particularly connected to the rationalist philosophers (e.g., Descartes and

Table 8.1 Kinship terms

Word	Semantemes
Father	male + parent
Mother	female + parent
Son	male + offspring
Daughter	female + offspring
Brother	male + sibling
Sister	female + sibling

Leibniz). The theory of semantic primitives is also related to “logical atomism” (Oliver, 1998), versions of which were put forward by Wittgenstein (1922) in his *Tractatus Logico-Philosophicus* and by Bertrand Russell (1924), both of whom were affiliated with logical positivism. (As is well known, Wittgenstein later changed his position and developed a more holistic and pragmatic view of language.) In linguistics, Chomsky has been the main representative of this rationalist strain of philosophy. Such a rationalist theory of semantics is similar to views put forward in IS, for example, in thesauri and in the facet-analytic tradition established by Ranganathan as well as in “formal concept analysis” (cf. Priss, 2006).

Although this rationalist theory dominates the literature (and is associated with the cognitive view), I do not find it fruitful for KO. First, the arguments that have been raised against it by the researchers mentioned here seem plausible. Second, semantic relations in KO are mostly a product of scientific ontological models; for example, the relations between chemical elements are not hardwired in our brains but are discovered by chemical researchers. Consequently, the creators of KOS have to identify the semantic relations in the subject literature rather than through psychological studies.

Concerning Semantics Specific to Given Empirical Languages

A paper by Hedlund et al. (2001) bears the title “Aspects of Swedish Morphology and Semantics from the Perspective of Mono- and Cross-Language Information Retrieval.” The wording of this title implies that the Swedish language has a semantics of its own. In other words, semantic relations are structural relations attributed to different empirical languages. This view is also evident in the literature of structural linguistics. As demonstrated in Table 8.2, the English word “tree” does not have the same meaning as the Danish word “træ.” Natural languages are structures in which the words classify the world differently.

Furthermore, many techniques in computational linguistics and natural language processing (NLP) are based on structures that are specific to a given language. For example, the commercial program Connexor (2003–2004, online) is described as giving

a semantic interpretation of the syntactic structure, which means that many language-specific patterns are normalized. For example, the Machine representation of the sentence “A book was given to John” shows the notional roles object and indirect object that correspond to the similar roles in “Somebody gave John a book.”

The focus on differences between different natural languages has been useful for IS. Research such as that by Hedlund et al. (2001) has provided knowledge that is very fruitful for IR. On the other hand, some

Table 8.2 Cultural relativity in word meanings

English	*German	*Danish	*French	Italian	Spanish
Tree	Baum	Træ skov	arbre	albero	Arbol
Wood	Holz		bois	legno	Leña
					Madera
Woods	Wald			bosco	Bosque
Forest			forêt	foresta	Selva

Originally presented by the Danish structural linguist Louis Hjelmslev (1943). Extended by information from Buckley (2001).

KOS (for example, the UDC) are applied across multiple languages and developed field by field. Semantic structures may be established in different domains and may diffuse into general languages. Our conceptions of uranium and radium as radioactive materials are based on scientific discoveries made within physics and transferred from there into general language. In other words, semantic structures in IS cannot be established simply by the study of natural languages: this also requires domain-specific knowledge.

Concerning Domain- or Discourse-Specific Semantics

As I noted earlier, pragmatism holds that descriptions and conceptions of objects are made from certain perspectives and involve certain pre-understandings and interests. This principle also figures prominently in other epistemological schemes, such as those of hermeneutics and Thomas Kuhn's theory of scientific paradigms. Although objects have objective properties, representation of those properties in languages and concepts is always more or less "subjective" or "biased" by individuals, social groups, or different cultures. Different human interests stress different properties of objects. Pharmacology and chemistry, for example, emphasize different properties of the same chemical elements: A chemical database emphasizes structural descriptions; a pharmacological database emphasizes medical effects.

The implication is that semantic relations reflect human interests. For example, pharmacology as a domain or discourse community emphasizes those semantic relations that are related to medical and side effects. This does not imply that all semantic relations are domain-specific. Pharmacology as a domain is heavily dependent on chemical research and the two domains share many concepts and semantic relations. Still, parts of their descriptions contain descriptions and semantic relations that reflect the specific goals of their respective domains.

How are the basic semantic structures determined within a domain? Keil (1989, p. 159) has outlined some important developments in theories about concepts and semantics:

The history of all natural sciences documents the discovery that certain entities that share immediate properties nonetheless belong to different kinds. Biology offers a great many examples, such as the discoveries that dolphins and whales are not fish but mammals, that the bat is not a kind of bird, that the glass “snake” is in fact a kind of lizard with only vestigial limbs beneath its skin. In the plant kingdom it has been found, for example, that some “vegetables” are really fruits and that some “leaves” are not really leaves. From the realm of minerals and elements have come the discoveries, among others, that mercury is a metal and that water is a compound.

In almost all these cases the discoveries follow a similar course. Certain entities are initially classified as members of a kind because they share many salient properties with other bona fide members of that kind and because their membership is in accordance with current theories. This classification may be accepted for centuries until some new insight leads to a realization that the entities share other, more fundamentally important properties with a different kind not with their apparent kind.

Sometimes it is discovered that although the fundamental properties of the entities are not those of their apparent kind, they do not seem to be those of any other familiar kind either. In such cases a new theoretical structure must develop that provides a meaningful system of classification.

There are many profound questions about when a discovery will have a major impact on a scheme of classification, but certainly a major factor is whether that discovery is made in the context of a coherent causal theory in which the discovered properties are not only meaningful but central.

This quotation shows that concepts and semantic structures depend on our worldviews and theories, including those shaped by scientific discoveries. It is also supportive of scientific realism, according to which science uncovers deeper and deeper layers of reality and in the process changes our theories, concepts, classification schemes, and semantics. Such a view is very different from the prevailing view that concepts are inherent in the mind or in specific languages.

In the literature of any domain, different theories and epistemologies come into play (cf. the lemma “domains” in Hjørland & Nicolaisen, 2005, online). In some cases (e.g., in psychology), different “schools” or “paradigms” co-exist, each with its own journal(s) (cf. Hjørland, 2002b). In most cases, however, such different epistemologies or paradigms are not self-conscious and do not have formally established information sources and communication structures. In the case of medicine, the movement known as *evidence-based medicine* may be considered a paradigm; but

there are no self-conscious alternative paradigms in medicine, a fact that challenges our view.⁶ In such cases, the existence of different paradigms has to be demonstrated by analyzing different methodologies and assumptions made in the field; studies of different paradigms (e.g., by using bibliometric methods) are much more difficult to perform. A working hypothesis is that different theories, background assumptions, and paradigms are at play in *any* field of knowledge (although, of course, the degree of consensus varies from field to field and variant views may be almost absent in some fields).

The meanings of particular words or symbols are primarily influenced by the dominant view or paradigm within a given domain or discourse. Any attempt to change the dominant view implies a need to reconsider established meanings. This is often not clear to the users of those words and symbols, who may use terms and symbols with meanings that work against what they are trying to do. When the need to redefine symbols has become clear to users, they may choose to use a different term or to continue to use a term with a somewhat different meaning. In this way, meanings are linked to different views, interests, and goals; accordingly, terms can generally be considered polysemous.⁷ Attempts to standardize terminology may unwittingly suppress certain views. This problem is, for example, important to consider in relation to The Unified Medical Language System (UMLS) project. Campbell, Oliver, Spackman, and Shortliffe (1998, pp. 426–427) have discussed how the UMLS has integrated the concept “Aspirin” from two different thesaural sources:

It is obvious that the intension associated with a term in a source terminology is represented at least in part by its location in a hierarchy and by decisions made regarding synonyms and non-synonyms. Aspirin in the CRISP Thesaurus is a chemical; it is also a centrally acting drug that has antirheumatic, anti-inflammatory, analgesic, and antipyretic properties. Similarly, the UMLS equivalent of aspirin in SNOMED, acetylsalicylic acid, is a chemical. It is also a drug with several of the same properties that it has in the CRISP Thesaurus: It is a centrally acting agent, an analgesic, and an antipyretic. On the other hand, in SNOMED, acetylsalicylic acid is not synonymous with two other UMLS equivalents of aspirin, Easprin and Zorprin, because the first is a generic drug and the other two are proprietary drugs. Thus, in SNOMED, the intension of aspirin is clearly not the same as the intension of Easprin, yet aspirin and Easprin are linked to the same CUI. It may even be argued that there are subtle differences in the intension of aspirin in CRISP and SNOMED, yet these differences are obscured or lost when one moves from the source terminology to the CUI.

How a term like “aspirin” should be defined and which semantic relations should be assigned in a given KOS is thus not an objective fact but a question related to the purpose of that KOS. As Campbell et al. (1998, p. 430) write:

In our previous discussion of how the UMLS represents “Aspirin,” ... we noted that most clinicians would probably not consider these three concepts [aspirin, Aspergum, and Ecotrin] interchangeable in the prescriptions they write. However, we also assert that from some possible perspectives, such as when we are concerned primarily with medication allergies, having these concepts all linked to the same extension makes perfect sense.

In this way, semantic decisions (such as whether aspirin, Aspergum, and Ecotrin should be considered synonymous terms) have to be decided by considering the consequences, such as whether these substances can be substituted for each other for the purpose that the KOS is designed to accomplish.

The implication of different paradigms for KO and semantics is that any bibliography of a certain size must confront conflicting ways of defining concepts and determining semantic relations. Literary warrant does not mean identifying only one text from which semantic relations may be inferred. The task is to negotiate between different claims put forward in different texts and to select the one that has the highest degree of cognitive authority or is considered best in relation to the goal of the KOS. Information scientists engaged in developing a given KOS have to negotiate among different views more or less visible in the literature to be indexed. In practice, this is often not done. The DDC, for example, claims to be based on the principle of literary warrant (Mitchell, 2001, p. 217); however, as Miksa (1994, p. 149) has noted, its practice has typically involved

arranging as many categories as possible in orders that reflected some kind of consensus among experts but thereafter simply doing something “practical” with the remainder. This appears to have been an approach characteristic of the DDC and the UDC as they developed over the years.

Systems such as the DDC are conservative because it is not economical to conduct deep literary investigations; to change the system; and, in particular, to reclassify books. Systems of this kind have to weigh the advantages of being updated in terms of literary warrant against the benefits of being a standard that is changed only rarely and reluctantly. There is a trade-off between being an optimal tool for the information seeker and a practical tool for the library manager. For the theory of IS, it is nonetheless important to describe the principles of designing optimal

search tools. Such principles have to deal with the conflicting criteria of literary warrant. For example, should *social psychology* be classified with psychology or with sociology? Bibliometric arguments might claim that as psychologists are dominant in social psychology, it should be classified with psychology. However, theoretical arguments might assert that the explanation of social psychological phenomena needs to be founded in sociological theory and so it should be classified with sociology. Historical and bibliometric studies have shown that there are actually two social psychologies—*psychological social psychology* (mainly experimental) and *sociological social psychology*. Each of these types of social psychology has its own courses, textbooks, journals, and so on, and so a third possibility would be to distinguish between psychological and sociological social psychology. The point is that the kind of information presented here is necessary for any informed decision about classification practice. Exactly the same kind of information would be helpful for the information seeker (in order to discriminate between the two kinds of social psychology or in order to find related information). If a semantic tool is to be optimized as a retrieval tool, such information about conflicting views of semantic relations should be available. This implies that classification research would make such alternatives visible in the literature and that the construction of systems would be based on such knowledge, with explicit references to, and interpretation of, literary warrant. The more that is invested in designing classification systems, the greater the benefits to the user. Arbitrary, easy, standardized, or “practical” solutions from an administrative point of view do not provide the information seeker with insights into the structures of knowledge.

The existence of different paradigms thus implies that any existing KOS can be examined in relation to both dominant and alternative views. As Ørom (2003) has demonstrated, different KOS such as the UDC and the DDC are more or less biased toward different paradigms within, for example, art studies. Although some systems (e.g., the *Art and Architecture Thesaurus* [Petersen, 1994]) are easier to adapt to new tendencies, there are no neutral platforms or criteria on which to base classifications and semantic tools. Any semantic tool may be more or less in harmony, or in conflict, with the views represented in the literature. Which view should the designer choose? The majority view? It is not possible to prescribe any single “correct” view or method for selecting a particular one. If it were, then it would be possible to prescribe how to do science, something that most philosophers of science find impossible. All we can conclude is that a precondition for designing quality KOS is that the designer knows the different views and is able to provide a reasonably informed and negotiated solution. In addition, the designer of a given KOS should analyze, from a pragmatic point of view, what goals the KOS seeks to fulfill.

Information scientists should ask the pragmatic question: Given the different interests and paradigms in the field, what kinds of interest should this specific system support? What difference does it make

whether some kinds of semantic relations are used at the expense of others? Perhaps the most important task of the information professional is to make the different interests and paradigms visible so that the user can make an informed choice.

Other Kinds of Warrant

In KO, as well as in IS in general, user-oriented and cognitive theories have flourished for some time. What kinds of “user warrant” exist with regard to semantic relations? Beghtol (1986) has discussed the following:

- Literary warrant and terminological warrant
- Scientific/philosophical warrant
- Educational warrant
- Cultural warrant

She does not, however, discuss user warrant. Indeed, it is difficult to imagine that the establishing of relations between terms A and B should be determined by investigating non-specialist users’ perspectives (e.g., that the classification of whales as mammals should be determined by users rather than by experts). In the case of popular music (Abrahamsen, 2003), the experts on genre are generally not the musicologists because so few of them have specialized in this field. It is closer to the users’ own expertise; however, journalists are presumably among those defining and naming new genres (and thus determining meaning and semantics). Other kinds of warrant may exist. Albrechtsen and Mark Pejtersen (2003) have argued for the existence of a sort of *work domain* warrant. This view may represent a tendency to prefer oral sources to written sources in IS. Yet, oral and written sources need the same kind of interpretation and argumentation. Information scientists may feel safer if they rely on “experts” rather than documents, but relevant documents are written by experts and are equally valid sources, if not more so.

Semantic Relations

Semantic relations are the relations between concepts, meanings, or senses. The concept [school] should be distinguished from the word “school.” [School] is a kind of [educational institution]. This is an example of a hyponymous, or hierarchical, relationship between two concepts or meanings, which is one among many kinds of semantic relations.

The concept [school] may, for example, be expressed by the terms or expressions “school,” “schoolhouse,” and “place for teaching.” The relation between “school” and “schoolhouse” is one of synonymy between two words, but the relation between “school” and “place for teaching” is a relation between a word and an expression. The relations between words are termed lexical relations.⁸ “School” also means [a group of people who

share a common outlook in relation to something] (as in “a school of thought”). This is a homonym relation: Two senses share the same word or expression—“school.” Synonyms and homonyms are not relations between concepts but are about concepts expressed with identical or with different signs.

Relations between concepts, senses, or meanings should not be confused with relations between the terms, words, expressions, or signs that are used to express the concepts. It is, however, common to mix both of these kinds of relations under the heading “semantic relations” (e.g., Cruse, 1986; Lyons, 1977; Malmkjær, 1995; Murphy, 2003). For this reason, synonyms, homonyms, and so forth, are considered under the label “semantic relations” in this chapter.

How many kinds of semantic relations exist? Is the number of semantic relations finite or infinite? What determines this number? Rosario and Hearst (2001) have observed that there are contradictory views in theoretical linguistics regarding the semantic properties of noun compounds (NCs). Some researchers hold that there exists a small set of semantic relationships that NCs may imply. Others maintain that the semantics of NCs cannot be exhausted by any finite listing of relationships. Green (2001, pp. 5–6) has argued that the inventory of semantic relationships includes both a closed set of relationships (including mainly hierarchical and equivalence relationships) and an open set of relationships. For example, every time a new verb is coined, the potential for the introduction of a new conceptual relationship arises.

Is it possible to draw up an exhaustive list of semantic relations? The answer is probably that any relation between objects (or processes or anything else) may be expressed in language because languages do not contain a limited number of semantic relations. “Love” is a relation between specific people, for example, Tom and Clare. [Tom] and [Clare] are thus individual concepts conjoined through the semantic relation “love.”⁹ (Note that the words “Tom” and “Clare” need not refer to the [Tom] and [Clare] in question, but may also refer to other individual concepts that do not share the same semantic relations.) The limit to the number of semantic relations seems to be relations that nobody has found interesting enough to conceptualize. If this argument is correct, then the number of semantic relations is infinite.

Different domains probably develop new kinds of semantic relations continuously. Rosario and Hearst (2001, pp. 83–84) identified 38 semantic relations within medicine.¹⁰

In this work we aim for a representation that is intermediate in generality between standard case roles (such as Agent, Patient, Topic, Instrument), and the specificity required for information extraction. We have created a set of relations that are sufficiently general to cover a significant number of noun compounds, but that can be domain specific enough to be useful in analysis. We want to support relationships between

entities that are shown to be important in cognitive linguistics, in particular we intend to support the kinds of inferences that arise from Talmy's force dynamics (Talmy, 1985). It has been shown that relations of this kind can be combined in order to determine the "directionality" of a sentence (e.g., whether or not a politician is in favor of, or opposed to, a proposal) (Hearst, 1990). In the medical domain this translates to, for example, mapping a sentence into a representation showing that a chemical removes an entity that is blocking the passage of a fluid through a channel.

The problem remains of determining what the appropriate kinds of relations are. In theoretical linguistics, there are contradictory views regarding the semantic properties of noun compounds (NCs). Levi (1978) argues that there exists a small set of semantic relationships that NCs may imply. Downing (1977) argues that the semantics of NCs cannot be exhausted by any finite listing of relationships. Between these two extremes lies Warren's (1978) taxonomy of six major semantic relations organized into a hierarchical structure.

We have identified the 38 relations shown in Table 1 [omitted here]. We tried to produce relations that correspond to the linguistic theories such as those of Levi and Warren, but in many cases these are inappropriate. Levi's classes are too general for our purposes; for example, she collapses the "location" and "time" relationships into one single class "In" and therefore field mouse and autumnal rain belong to the same class. Warren's classification schema is much more detailed, and there is some overlap between the top levels of Warren's hierarchy and our set of relations.

Rosario and Hearst (2001) thus seem to support the view that the number of semantic relations is infinite. In this regard, it is worth noting that semantic relations resemble commonly used grammatical categories. Now, categories and grammatical relations represent abstractions. Thus, our earlier example of a semantic relation, "love," may be seen as a special case of "being affected" (one of Aristotle's categories). Although the number of semantic relations appears to be unlimited, only a limited number of generalized relations tend to be used in practice.

In IR, the basic function of semantic relations is to contribute to the increase of recall and precision. For example, the inclusion of synonyms and broader terms in a query may contribute to increased recall, whereas the differentiation of homonyms and the specification of terms may increase precision. In this way, the wide use of the standard semantic relations employed in thesauri may be explained functionally. There are, however, recommendations that the number of relations should be expanded:

The participants [in a NISO 1999 workshop on standards for electronic thesauri] recommended that a much richer, hierarchically organized, set of relationships be developed. ... There is reason to expect that provision for semantic relations in controlled vocabularies will become much more extensive in a future standard. (Milstead, 2001, p. 65)

How should we explain this demand for a much richer set of relationships than that ordinarily used in, for example, thesauri? The answer may imply a criticism of the traditional recall/precision way of understanding IR. What information searchers need are maps that inform them about the world (and the literature about that world) in which they live and act. They need such maps in order to formulate questions in the first instance. In order to formulate queries and to interact with information sources, advanced semantic tools are often very useful. This is probably especially so in the humanities, where concepts are more clearly associated with worldviews. The notion of conceptual history (*Begriffsgeschichte*) as developed in Germany provides a good illustration of this point. Historians and other humanistic researchers have realized that in order to use sources from a given period, one must know what the terms meant at the time. Therefore, they have developed impressive historical dictionaries that provide detailed information about conceptual developments within different domains, just as they have developed methodological principles on how to work with historical information sources (cf. Hampsher-Monk, Tilmans, & van Vree, 1998).

An example of a semantic tool developed in this tradition is *Reallexikon der deutschen Literaturwissenschaft* (Weimar, 1997–2003), which provides the following information for each term:

- The term (e.g., “bibliography”)
- A definition (e.g., definition of “bibliography”)
- A history (i.e., etymology) of the word (e.g., the etymology of the word “bibliography”)
- A history of the concept (e.g., the history of the meanings of “bibliography”)
- A history of the field (e.g., the history of bibliographies themselves)
- A history of research about the field (e.g., the history of research on bibliographies, i.e., library science)

I mention this example because it illustrates the existence of important work that may inspire IS to adopt a broader approach to semantic relations. To date, few researchers have investigated whether different domains need different kinds of semantic tools displaying different kinds of semantic relations: A notable exception is Roberts (1985), who

has argued for the importance of specific kinds of relations in the social sciences.

The “Intellectual” Versus the Social Organization of Knowledge

Are there semantic relations between citing papers and their cited papers? Some authors have explicitly used this terminology (e.g., Harter, Nisonger, & Weng, 1993; Qin, 1999; Song & Galardi, 2001). Others have used bibliometric methods in order to establish semantic relations in thesauri and information retrieval (e.g., Kessler, 1965; Pao, 1993; Rees-Potter, 1989, 1991; Salton, 1971; Schneider, 2004), thus implying such a relation.

Harter et al. (1993) examined semantic relations between citing and cited papers by applying two methods: a macro analysis, based on a comparison of the Library of Congress class numbers assigned citing and cited documents, and a microanalysis, based on a comparison of descriptors assigned to citing and cited documents by three indexing and abstracting services, *ERIC*, *LISA*, and *Library Literature*. Both analyses suggested that the subject similarity among pairs of cited and citing documents is typically very small (at least in this domain). In interpreting the results of this study, one should remember that subject determination typically is a process with great uncertainty and variance. If two documents, A and B, have a citing relation (directly or indirectly by co-citations or bibliographic coupling), they might be understood as semantically related whether or not they are assigned the same descriptors or classification codes by somebody (or whether or not they contain the same words, for that matter: one might, for example, be in English, the other in Danish). I hold that the citing relation is in itself a kind of semantic relation. In support of this claim, I distinguish between “ontological” and social semantic relations and argue that citing relations belong to the latter.

The kinds of relations typically used in semantic tools are “real” relations such as geographical relations (e.g., Denmark is part of Europe), biological relations (e.g., cats are mammals), and chemical relations (such as the relations implied by the periodic table). Such relations are “ontological.” Researchers produce ontological models that are used to organize knowledge.

A “social relation” is a different kind of relation. For example, disciplinary relations are social. The classification of sociology as a social science means that sociologists belong to the community of social scientists. A discipline is a social concept defined as people with similar education or other social ties, such as sharing the same organizations and journals. Disciplines typically have strong internal citation relations in comparison to their relations to other disciplines. A citation network is thus a kind of social relationship.

In some cases, ontological models of reality correspond very well with social organizations such as disciplines or citation networks. In other cases, the connections may be weak (many disciplines or “schools” may, for example, have overlapping ontological structures). Social constructivists tend to claim that ontological models and discoveries are just constructed: In other words, the social organization of knowledge is somehow primary to the intellectual organization. Scientific realists, on the other hand, tend to see ontological structures as primary and social structures as based on preexisting structures discovered by science.

Ontological models and theories developed by researchers as well as social organizations provide meaning to terms and semantic relations between terms. One may discuss which kind of meanings or relations are the most truthful or fruitful ones. However, information scientists provide semantic tools that are based on both kinds of relations. Bibliometric tools and tools based on ontological relations are available and in many cases supplement each other in IR. One should study the ways in which they supplement each other. In other words, semantic relations as provided by citing relations are legitimate in their own right. There is no need to verify them as Harter et al. (1993) and Schneider (2004) attempt to do. A traditional thesaurus and a bibliometric map may, in different ways, inform a person seeking information. Their relative value may depend on domain-specific issues such as how terminology is used and whether citation patterns reflect relevant specializations. A citation relation between two papers, A and B, is in itself a semantic relation, regardless of whether it corresponds with how A and B are otherwise determined to be related.

Conclusion

The pragmatist view of semantics suggests that words and expressions are tools for interaction and their meanings are their functions within the interaction, constituting their capacities to serve it in their distinctive ways. When information professionals classify documents or informational objects, the relevant meanings and properties are available only on the basis of some descriptions. This important consideration, which van Rijsbergen (1979) has emphasized, stands in opposition to the prevailing implicit assumption that all relevant properties of the objects are obvious to information specialists and that the latter follow certain given principles providing an optimal classification that is objective, neutral, and universal—hence, technically efficient. Hunter’s (2002, p. 25) textbook on classification demonstrates how machine bolts may be classified according to their material, thread size, head shape, and finish. Admittedly, this example is probably not typical of documentary classification (it is classification made too simple). The same thing is often described differently for different purposes. Differing human interests emphasize different properties of objects. A typical database,

on which IR experiments are performed, is best conceptualized as a merging of different descriptions serving different purposes.

Traditional approaches to KO have a tighter affiliation with positivism than with the pragmatist view of semantics. The solutions provided have not been based on the view that a typical database, on which IR experiments are performed, should be conceived of as a merging of different descriptions serving different purposes and based on different epistemologies. The implication is that traditional views have provided solutions that are, at best, statistical averages and thus sub-optimal. The prospect of KO based on a pragmatist understanding of semantics holds open the promise of fine-tuning KOS in different domains and genres.

Endnotes

1. LIS and IS are regarded as synonyms in this chapter. Other researchers do not regard them as synonyms. This example of semantic relations is an illustration of the problems that KO faces. Those who claim that the two terms are not synonyms should be able to say whether a given paper belongs to IS or to LIS.
2. In the sociology of science, the debate is between “meaning finitism” and “meaning determinism,” a related theoretical discussion (cf. Barnes, 2002; Bloor, 1997, pp. 1–3, 9–11; Haukioja, 2005; Klaes, 2002; Larsson, 2003; and Weber, 2005). Harris (2005) provides an important critique of the semantic assumptions generally made in science.
3. Some texts define semantic relations as stable and different from “syntactic relations” (Foskett, 1977, p. 72) or from pragmatic relations (Dahllöf, 1999, p. 44). Such positions are not in accordance with the theoretical view put forward in this chapter and would make the question “Under what conditions can a semantic relation be said to exist?” meaningless.
4. Sowa (2000, online) writes about Ogden & Richards’s (1923) triangle of meaning: “The triangle in Figure [8.]1 has a long history. Aristotle distinguished objects, the words that refer to them, and the corresponding experiences in the *psychê*. Frege and Peirce adopted that three-way distinction from Aristotle and used it as the semantic foundation for their systems of logic. Frege’s terms for the three vertices of the triangle were *Zeichen* (sign) for the symbol, *Sinn* (sense) for the concept, and *Bedeutung* (reference) for the object.”
5. Regarding relativism in color concepts see in addition to Ratner (1989) also Goodwin (2000), Lucy (1997), Roberson, Davies, and Davidoff (2000), and Saunders (1998).
6. Perhaps “narrative based medicine” (Greenhalgh & Hurwitz, 1998) should be considered a competing paradigm.
7. This is clearly seen in the German tradition of *Begriffsgeschichte*, which is discussed in the section on semantic relations.

8. "Lexical Semantics is about the meaning of words. Although obviously a central concern of linguistics, the semantic behaviour of words has been unduly neglected in the current literature, which has tended to emphasize sentential semantics and its relation to formal systems of logic" (Cruse, 1986, book cover).
9. Such relations could be drawn, for example, in semantic networks. See figure 7 in McCann (1997).
10. Rosario and Hearst (2004) described the problems involved in distinguishing seven relation types between the entities "treatment" and "disease" in biomedical texts.

References

- Abrahamsen, K. T. (2003). Indexing of musical genres: An epistemological perspective. *Knowledge Organization*, 30, 144–169.
- Albrechtsen, H., & Mark Pejtersen, A. (2003). Cognitive work analysis and work centered design of classification schemes. *Knowledge Organization*, 30, 213–227.
- Antoniou, G., & van Harmelen, F. (2004). *A semantic Web primer*. Cambridge, MA: MIT Press.
- Bada, M., Stevens, R., Goble, C., Gil, Y., Ashburner, M., Blake, J. A., et al. (2004). A short study on the success of the gene ontology. *Journal of Web Semantics*, 1, 235–240. Retrieved December 15, 2005, from www.websemanticsjournal.org/ps/pub/2004-9
- Barnes, B. (2002). Thomas Kuhn and the problem of social order in science. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 122–141). Cambridge, UK: Cambridge University Press.
- Bean, C. A., & Green, R. (Eds.). (2001). *Relationships in the organization of knowledge*. Dordrecht, NL: Kluwer.
- Beghtol, C. (1986). Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, 30, 109–125.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. *Scientific American*, 284(5), 34–43.
- Blair, D. C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier.
- Blair, D. C. (2003). Information retrieval and the philosophy of language. *Annual Review of Information Science and Technology*, 37, 3–50.
- Bliss, H. E. (1929). *The organization of knowledge and the system of the sciences*. New York: Henry Holt.
- Bliss, H. E. (1935). *A system of bibliographical classification*. New York: H. W. Wilson.
- Bloomfield, L. (1936/1985). Language or ideas? In J. J. Katz (Ed.), *The philosophy of linguistics* (pp. 19–25). Oxford, UK: Oxford University Press.
- Bloor, D. (1997). *Wittgenstein: Rules and institutions*. London: Routledge.
- Bonnevie, E. (2001). Dretske's semantic information theory and meta-theories in library and information science. *Journal of Documentation*, 57, 519–534.

- Brooks, T. A. (1995). Topical subject expertise and the semantic distance model of relevance assessment. *Journal of Documentation*, 51, 370–387.
- Brooks, T. A. (1998). The semantic distance model of relevance assessment. *Proceedings of the Annual Meeting of the American Society for Information Science*, 33–44.
- Broughton, V., Hansson, J., Hjørland, B., & López-Huertas, M. J. (2005). Knowledge organisation: Report of working group 7. In L. Kajberg & L. Lørring (Eds.), *European Curriculum Reflections on Education in Library and Information Science*. Copenhagen: Royal School of Library and Information Science. Retrieved December 15, 2005, from www.db.dk/LIS-EU/workshop.asp
- Buckley, G. (2001). *Semantics*. Retrieved July 31, 2005, from www.ling.upenn.edu/courses/Spring_2001/ling001/semantics.html
- Budd, J. M. (2004). Relevance: Language, semantics, philosophy. *Library Trends*, 52, 447–462.
- Campbell, K. E., Oliver, D. E., Spackman, K. A., & Shortliffe, E. H. (1998). Representing thoughts, words, and things in the UMLS. *Journal of the American Medical Informatics Association*, 5, 421–431. Retrieved December 15, 2005, from www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61323
- Carroll, L. (1899). *Through the looking glass*. New York: M. F. Mansfield & A. Wessels.
- Chaffin, R., Herrmann, D. J., & Winston, M. (1988). A taxonomy of part-whole relations. *Cognition and Language*, 3, 1–32.
- Connexor. (2003–2004). *Machine semantics*. Retrieved December 15, 2005, from www.connexor.com/software/semantics
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.
- Cruse, D. A. (2001). Lexical semantics. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (Vol. 13, pp. 8758–8764). Amsterdam: Elsevier.
- Dahlberg, I. (1978). A referent-oriented, analytical concept theory for INTERCONCEPT. *International Classification*, 5, 142–151.
- Dahlberg, I. (1995). Conceptual structures and systematization. *IFID Journal*, 20(3), 9–24.
- Dahllöf, M. (1999). *Språklig betydelse: En introduktion till semantik och pragmatik* [Linguistic meaning: An introduction to semantics and pragmatics]. Lund: Studentlitteratur.
- Daily, J. E. (1979). Semantics. In A. Kent, H. Lancour, & J. E. Daily (Eds.), *Encyclopedia of Library and Information Science* (Vol. 27, pp. 209–215). New York: Dekker.
- Ding, C. H. Q. (2005). A probabilistic model for latent semantic indexing. *Journal of the American Society for Information Science and Technology*, 56, 597–608.
- Doerr, M. (2001). Semantic problems of thesaurus mapping. *Journal of Digital Information*, 1(8). Retrieved December 15, 2005, from <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr>
- Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 53, 810–842.
- Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38, 189–230.
- Ereshefsky, M. (2000). *The poverty of the Linnaean hierarchy: A philosophical study of biological taxonomy*. Cambridge, UK: Cambridge University Press.

- Fensel, D., Hendler, J. A., Lieberman, H., & Wahlster, W. (Eds.). (2003). *Spinning the semantic Web: Bringing the World Wide Web to its full potential*. Cambridge, MA: MIT Press.
- Foskett, A. C. (1977). Assigned Indexing I: Semantics. In *The subject approach to information* (pp. 67–85). London: Clive Bingley.
- Frohmann, B. P. (1983). An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG. *Cataloging & Classification Quarterly*, 4, 11–27.
- Goodwin, C. (2000). Practices of color classification. *Mind, Culture and Activity*, 7, 19–36.
- Green, R. (2001). Relationships in the organization of knowledge: An overview. In C. A. Bean & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 3–18). Dordrecht, NL: Kluwer.
- Green, R., Bean, C. A., & Myaeng, S. H. (Eds.). (2002). *The semantics of relationships: An interdisciplinary perspective*. Dordrecht, NL: Kluwer Academic Publishers.
- Greenhalgh, T., & Hurwitz, B. (1998). *Narrative based medicine: Dialogue and discourse in clinical practice*. London: BMJ Publishing Group.
- Hammwöhner, R., & Kuhlen, R. (1994). Semantic control of open hypertext systems by typed objects. *Journal of Information Science*, 20, 175–184.
- Hampsher-Monk, I., Tilmans, K., & van Vree, F. (Eds.). (1998). *History of concepts: Comparative perspectives*. Amsterdam: Amsterdam University Press.
- Harris, R. (2005). *The semantics of science*. London: Continuum International Publishing Group.
- Harter, S. P., Nisonger, T. E., & Weng, A. W. (1993). Semantic relations between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science*, 44, 543–552.
- Haukioja, J. (2005). A middle position between meaning finitism and meaning Platonism. *International Journal of Philosophical Studies*, 13, 35–51.
- Hearst, M. A. (1990). A hybrid approach to restricted text interpretation. In P. S. Jacobs, (Ed.), *Text-based intelligent systems: Current research in text analysis, information extraction and retrieval* (pp. 38–43). Schenectady, NY: GE Research & Development Center.
- Hedlund, T., Pirkola, A., & Kalervo, J. (2001). Aspects of Swedish morphology and semantics from the perspective of mono- and cross-language information retrieval. *Information Processing & Management*, 37, 147–161.
- Hjelmslev, L. (1943). *Omkring sprogteoriens grundlæggelse*. Copenhagen: B. Lunos bogtrykkeri a/s.
- Hjørland, B. (1997). *Information seeking and subject representation: An activity-theoretical approach to information science*. Westport, CT: Greenwood.
- Hjørland, B. (1998). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25, 16–31. Retrieved December 15, 2005, from www.db.dk/bh/publikationer/Filer/ir_semant_2.pdf
- Hjørland, B. (2002a). Domain analysis in information science: Eleven approaches—traditional as well as innovative. *Journal of Documentation*, 58, 422–462. Retrieved December 15, 2005, from www.db.dk/bh/publikationer/Filer/JDOC_2002_Eleven_approaches.pdf

- Hjørland, B. (2002b). Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology*, 53, 257–270.
- Hjørland, B., & Nicolaisen, J. (2005). *The epistemological lifeboat*. Copenhagen: Royal School of Library and Information Science. Retrieved December 15, 2005, from www.db.dk/jni/lifeboat/home.htm
- Hjørland, B., & Nissen Pedersen, K. (2005). A substantive theory of classification for information retrieval. *Journal of Documentation*, 61, 582–597. Retrieved December 15, 2005, from www.db.dk/bh/Core%20Concepts%20in%20LIS/Hjorland%20&%20Nissen.pdf
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: Beyond traditional authority files* (CLIR Report 91). Washington, DC: Council on Library and Information Resources. Retrieved December 15, 2005, from www.clir.org/pubs/reports/pub91/pub91.pdf
- Hunter, E. J. (2002). *Classification made simple* (2nd ed.). Aldershot, UK: Ashgate.
- Iris, M. A., Litowitz, B., & Evens, M. (1988). Problems of the part-whole relation. In M. W. Evens (Ed.), *Relational models of the lexicon* (pp. 261–288). Cambridge, UK: Cambridge University Press.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16, 223–233.
- Khoo, C., & Na, J.-C. (2006). Semantic relations in information science. *Annual Review of Information Science and Technology*, 40, 157–228.
- Kinkade, R. G. (Ed.). (1974). *Thesaurus of psychological index terms*. Washington, DC: American Psychological Association.
- Klaes, M. (2002). Some remarks on the place of psychological and social elements in a theory of custom. *American Journal of Economics and Sociology*, 61, 519–530. Retrieved December 15, 2005, from www.findarticles.com/p/articles/mi_m0254/is_2_61/ai_86469072#continue
- Larsson, J. (2003). *Finitism and symmetry: An inquiry into the basic notions of the strong programme*. Unpublished doctoral dissertation, Göteborg University, Sweden.
- Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
- Longman synonym dictionary. (1986). Essex, UK: Longman.
- Lucy, J. (1997). The linguistics of “color.” In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 320–346). Cambridge, UK: Cambridge University Press.
- Lyons, J. (1977). *Semantics*. Cambridge, UK: Cambridge University Press.
- McCann, J. M. (1997). *Generation of marketing insights: Semantic networks*. Retrieved December 15, 2005, from <http://web.archive.org/web/19990127092407/http://www.duke.edu/~mccann/mwb/15semnet.htm>
- Malmkjær, K. (1995). Semantics. In K. Malmkjær (Ed.), *The linguistics encyclopedia* (pp. 389–398). London: Routledge.
- Miksa, F. (1994). Classification. In W. A. Wiegand & D. G. Davis (Eds.), *Encyclopedia of library history* (pp. 144–153). New York: Garland Publishing.
- Miller, G. A. (1998). Nouns in WordNet. In C. Felbaum (Ed.), *WordNet: An electronic lexical database* (pp. 23–46). Cambridge, MA: MIT Press.

- Milstead, J. L. (2001). Standards for relationships between subject indexing terms. In C. A. Bean & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 53–66). Dordrecht, NL: Kluwer.
- Mitchell, J. S. (2001). Relationships in the Dewey Decimal Classification System. In C. A. Bean & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 211–226). Dordrecht, NL: Kluwer.
- Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy, and other paradigms*. Cambridge, UK: Cambridge University Press.
- National Library of Medicine. (2005). *Frequently asked questions: Who are the indexers, and what are their qualifications?* Retrieved December 15, 2005, from www.nlm.nih.gov/bsd/indexfaq.html#qualifications
- Nida, E. A. (1975). *Componential analysis of meaning: An introduction to semantic structures*. The Hague, NL: Mouton.
- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning: A study of the influence of language upon thought and of the science of symbolism*. London: Routledge & Kegan Paul.
- Oliver, A. (1998). Logical atomism. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (Vol. 5, pp. 772–775). London: Routledge.
- Ørom, A. (2003). Knowledge organization in the domain of art studies: History, transition and conceptual changes. *Knowledge Organization*, 30, 128–143.
- Pao, M. L. (1993). Term and citation retrieval: A field study. *Information Processing & Management*, 29, 95–112.
- Peirce, C. S. (1931–1958). *Collected papers of C. S. Peirce* (C. Hartshorne, P. Weiss, & A. Burks, Eds.). Cambridge, MA: Harvard University Press.
- Peregrin, J. (2004). Pragmatism und Semantik [Pragmatism and semantics]. In A. Fuhrmann & E. J. Olsson (Eds.), *Pragmatisch denken* [Thinking pragmatically] (pp. 89–108). Frankfurt am Main, Germany: Ontos. English manuscript version retrieved December 15, 2005, from <http://jarda.peregrin.cz/mybibl/PDFTxt/482.pdf>
- Petersen, T. (Ed.). (1994). *Art and architecture thesaurus* (2nd ed.). New York: Oxford University Press.
- Priss, U. (2006). Formal concept analysis in information science. *Annual Review of Information Science and Technology*, 40, 521–543.
- Putnam, H. (1975). The meaning of “meaning.” In K. Gunderson (Ed.), *Language, mind, and knowledge* (pp. 131–193). Minneapolis: University of Minnesota Press.
- Qin, J. (1999). Discovering semantic patterns in bibliographically coupled documents. *Library Trends*, 48, 109–132.
- Qin, J. (2000). Semantic similarities between a keyword database and a controlled vocabulary database: An investigation in the antibiotic resistance literature. *Journal of the American Society for Information Science*, 51, 166–180.
- Ratner, C. (1989). A sociohistorical critique of naturalistic theories of color perception. *Journal of Mind and Behavior*, 10, 361–372. Retrieved December 15, 2005, from <http://web.archive.org/web/20031029152929/http://www.humboldt1.com/~cr2/colors.htm>
- Read, C. S. (1973). General semantics. In A. Kent, H. Lancour, & J. E. Daily (Eds.), *Encyclopedia of library and information science* (Vol. 9, pp. 211–221). New York: Dekker.

- Rees-Potter, L. K. (1989). Dynamic thesaural systems: A bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management*, 25, 677–691.
- Rees-Potter, L. K. (1991). Dynamic thesauri: The cognitive function. *Proceedings of the 1st International ISKO Conference, Darmstadt, Part 2*, 145–150.
- Roberts, N. (1985). Concepts, structures and retrieval in the social sciences up to c. 1970. *Social Science Information Studies*, 5, 55–67.
- Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129, 369–398.
- Rosario, B., & Hearst, M. (2001). Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, 82–90. Retrieved December 15, 2005, from <http://biotext.berkeley.edu/papers/emnlp01.pdf>
- Rosario, B., & Hearst, M. (2004). Classifying semantic relations in bioscience texts. *42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, 430–437. Retrieved December 15, 2005, from <http://biotext.berkeley.edu/papers/acl04-relations.pdf>
- Russell, B. (1924). Logical atomism. In R. C. Marsh (Ed.), *Logic and knowledge* (pp. 323–343). London: Allen & Unwin.
- Salton, G. (1971). Automatic indexing using bibliographic citations. *Journal of Documentation*, 27, 98–110.
- Saunders, B. (1998, August). *Revisiting Basic Color Terms*. Paper presented at conference on “Anthropology and Psychology: The Legacy of the Torres Strait Expedition,” St. John's College, Cambridge. Retrieved January 31, 2006, from <http://human-nature.com/science-as-culture/saunders.html>
- Schneider, J. (2004). *Verification of bibliometric methods' applicability for thesaurus construction*. Unpublished doctoral dissertation, Royal School of Library and Information Science, Aalborg, Denmark. Retrieved December 15, 2005, from <http://biblis.db.dk/uhtbin/hyperion.exe/db.jessch04>
- Soergel, D. (1995). The Art and Architecture Thesaurus (AAT): A critical appraisal. *Visual Resources*, 10, 369–400. Retrieved December 15, 2005, from www.dsoergel.com/cv/B47_short.pdf (short version); www.dsoergel.com/cv/B47_long.pdf (long version)
- Song, M., & Galardi, P. (2001). Semantic relationships between highly cited articles and citing articles in information retrieval. *Proceedings of the Annual Meeting of the American Society for Information Science*, 171–181.
- Sowa, J. F. (2000). Ontology, metadata, and semiotics. In B. Ganter & G. W. Mineau (Eds.), *Conceptual structures: Logical, linguistic, and computational issues* (pp. 55–81). Berlin: Springer-Verlag. Retrieved December 15, 2005, from <http://users.bestweb.net/~sowa/peirce/ontometa.htm>
- Sparck Jones, K. (1992). Thesaurus. In S. C. Shapiro (Ed.), *Encyclopedia of artificial intelligence* (Vol. 2, pp. 1605–1613). New York: Wiley.
- Sparck Jones, K. (2005). Revisiting classification for retrieval. *Journal of Documentation*, 61, 598–601. Retrieved December 15, 2005, from www.db.dk/bh/Core%20Concepts%20in%20LIS/Sparck%20Jones_reply%20to%20Hjorland%20&%20Nissen.pdf

- Stokolova, N. A. (1976). Syntactic tools and semantic power of information languages (Pt. II of "Elements of a semantic theory of information retrieval"). *International Classification*, 3, 75–81.
- Stokolova, N. A. (1977a). Elements of a semantic theory of information retrieval—I. The concepts of relevance and information language. *Information Processing & Management*, 13, 227–234.
- Stokolova, N. A. (1977b). Paradigmatic relations (Pt. III of "Elements of a semantic theory of information retrieval"). *International Classification*, 4, 11–19.
- Talmy, L. (1985). Force dynamics in language and thought. In W. F. Eikfont, P. D. Kroeber, & K. L. Peterson (Eds.), *Papers from the Parasession on Causatives and Agentivity* (pp. 293–337). Chicago: Chicago Linguistic Society.
- Thesaurus of ERIC Descriptors* (11th ed.) (1987). Phoenix, AZ: Oryx Press.
- van Rijsbergen, C. J. (1979). *Information retrieval* (2nd ed.). London: Butterworths. Retrieved December 15, 2005, from www.dcs.gla.ac.uk/Keith/Chapter.3/Ch.3.html
- van Rijsbergen, C. J. (1986). A new theoretical framework for information retrieval. *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 194–200.
- Vickery, B. C., & Vickery, A. (1987). Semantics and retrieval. In *Information science in theory and practice* (pp. 133–179). London: Bowker-Saur.
- Walker, A. (Ed.). (1997). *Thesaurus of psychological index terms* (8th ed.). Washington, DC: American Psychological Association.
- Warren, B. (1978). *Semantic patterns of noun-noun compounds* (Gothenburg Studies in English, 41). Gothenburg, Sweden: Acta Universitatis Gothoburgensis.
- Weber, M. (2005, October). *How strong is the case for social relativism in science?* Lecture held at the Minnesota Center for Philosophy of Science. Retrieved December 15, 2005, from http://philosophy.duke.edu/pdf/weber_duke.pdf
- Weimar, K. (Ed.). (1997–2003). *Reallexikon der deutschen Literaturwissenschaft, Band 1–3*. (3. neubearb. Aufl.). Berlin: Walter de Gruyter.
- Wellisch, H. H. (2000). *Glossary of terminology in abstracting, classification, indexing, and thesaurus construction* (2nd ed.). Medford, NJ: Information Today.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, 49, 327–355.
- Wierzbicka, A. (1996). *Semantics: Primes and universals*. Oxford, UK: Oxford University Press.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul. Hypertext of the Ogden bilingual edition retrieved July 31, 2005, from www.kfs.org/~jonathan/witt/tlph.html
- WordNet 2.1. A lexical database for the English language*. (2005). Princeton, NJ: Princeton University Cognitive Science Laboratory. Retrieved December 15, 2005, from <http://wordnet.princeton.edu>

Appendix

Some important kinds of semantic relations that have been presented in the literature:

1. **Active relation:** A semantic relation between two concepts, one of which expresses the performance of an operation or process affecting the other. The inverse of the passive relation.
2. **Antonymy:** A semantic relation in which A is the opposite of B (e.g., cold is the opposite of hot).
3. **Associative relation:** A semantic relation defined psychologically as the mental association of concepts (i.e., A is mentally associated with B by somebody). Often, associative relations are simply unspecified relations. In thesauri, antonyms are not usually specified but may be listed along with terms representing other kinds of relations under "associative relations."
4. **Causal relation:** A semantic relation in which A is the cause of B (e.g., a lack of vitamin C causes scurvy).
5. **Homonymy:** A semantic relation in which two concepts, A and B, are expressed by the same symbol (e.g., both a financial institution and the edge of a river are expressed by the word "bank"; i.e., the word has two senses).
6. **Hyponymous relations (hyponym-hyperonym):** Relations in which A is a kind of B; A is subordinate to B; A is narrower than B; B is broader than A. Also known as generic relation, genus-species relation, or hierarchical subordinate relation.
7. **Is-a relation:** A semantic relation between a general concept and individual instances of that concept; that is, A is an example, or instance, of B (e.g., Copenhagen is an instance of the general concept "capital").
8. **Locative relation:** A relation in which a concept indicates a location of a thing designated by another concept: that is, A is located in B (e.g., minorities in Denmark).
9. **Paradigmatic relation:** As defined by Wellisch (2000, p. 50), "a semantic relation between two concepts, that is considered to be either fixed by nature, self-evident, or established by convention. Examples: mother/child; fat/obesity; a state/its capital city."
10. **Partitive (i.e., part-whole) relation (meronymy):** a relationship between the whole and its parts; that is, A is part of B. A meronym is the name of a constituent part of, the substance of, or a member of something. Meronymy is the opposite of holonymy (i.e., B has A as part of itself).
11. **Passive relation:** A semantic relation between two concepts, one of which is affected by, or subjected to, an operation or process expressed by the other. The inverse of the active relation.
12. **Polysemy:** A mode of semantic relation in which a word has several subsenses that are related with one another (i.e., concepts A1, A2, and A3 are all expressed by the word "A"). Such a word is termed "polysemous" or "polysemantic."

13. Possessive relation: a semantic relation between a possessor and what is possessed (i.e., A belongs to B; B possesses A).
14. Related term: A term that is semantically related to another term. In thesauri, related terms are often coded RT and used for kinds of semantic relations other than synonymy (USE, UF), homonymy (separated by parenthetical qualifier), and generic relations and/or partitive relations (BT, NT). Related terms may, for example, express antagonistic relations, active/passive relations, causal relations, locative relations, or paradigmatic relations.
15. Synonymy: A semantic relation in which A denotes the same as B; A is equivalent with B.
16. Temporal relation: A semantic relation in which a concept indicates a time or period of an event designated by another concept (e.g., Second World War, 1939–1945).
17. Troponymy: According to WordNet 2.1 (2005), “the semantic relation of being a manner of [doing] something.”